Learning models for video quality prediction over wireless local area network and universal mobile telecommunication system networks

A. Khan   L. Sun   E. Ifeachor

Centre for Signal Processing and Multimedia Communication, School of Computing and Mathematics, University of Plymouth, Plymouth PL4 8AA, UK
E-mail: asiya.khan@plymouth.ac.uk

Abstract: Universal mobile telecommunication system (UMTS) is a third-generation mobile communications system that supports wireless wideband multimedia applications. The primary aim of this study is to present learning models based on neural networks for objective, non-intrusive prediction of video quality over wireless local area network (WLAN) and UMTS networks for video applications. The contributions of this study are two-fold: first, an investigation of the impact of parameters both in the application and physical layer on end-to-end video quality is presented. The parameters considered in the application layer are content type (CT), sender bitrate (SBR) and frame rate (FR), whereas in the physical layer block error rate (BLER) and link bandwidth (LBW) are considered. Secondly, learning models based on adaptive neural fuzzy inference system (ANFIS) are developed to predict the visual quality in terms of the mean opinion score for all contents over access networks of UMTS and WLAN. ANFIS is well suited for video quality prediction over error-prone and bandwidth restricted UMTS as it combines the advantages of neural networks and fuzzy systems. The ANFIS-based artificial neural network is trained using a combination of physical layer parameters such as BLER and LBW and application layer parameters of CT, SBR and FR. The proposed models are validated using unseen data set. The preliminary results show that good prediction accuracy was obtained from the models. This study should help in the development of a reference-free video prediction model and quality of service control methods for video over UMTS/WLAN networks.

1 Introduction

Transmission of multimedia applications and services over wireless networks is gaining popularity. Video transmission for mobile terminals is likely to be a major application in future mobile systems and a key factor for their success. Universal mobile telecommunication system (UMTS) networks are capable of providing high mobility, whereas wireless local area networks (WLANs) are known for having relatively higher bandwidths. Therefore ubiquitous data services and relatively high data rates across heterogeneous networks could be achieved by interworking 3G cellular networks with WLANs. This will enable a user to access 3G cellular services via a WLAN, while roaming within a range of hotspots. Thus, WLANs can be considered as a complementary technology for the 3G cellular data networks as well as a compulsory element of the future next generation mobile network. Interworking between WLAN and other cellular mobile networks is under extensive research by international standardisation forums (e.g. European Telecommunication Standard Institute, 3GPP, UMTS forum, etc.) [1, 2].

Video streaming is a multimedia service, which is recently gaining popularity and expected to unlock new revenue flows for mobile network operators. However, for such services to be successful, user's perceived quality of service (QoS) or the quality of experience (QoE) is likely to be the major determining factor. There are several factors that affect the QoE of multimedia applications and can be classified as

factors in the application and physical layers. In the application layer, QoE is driven by factors such as resolution, frame rate (FR), sender bitrate (SBR), video codec type, and so on. In the physical layer impairments such as the block error rate (BLER), jitter, delay, latency, and so on are introduced. Video quality can be evaluated either subjectively or based on objective parameters. Subjective quality is the users' perception of service quality (ITU-T P.800) [3]. The most widely used metric is the mean opinion score (MOS). Although subjective quality is the most reliable method, it is time-consuming and expensive and hence, the need for an objective method that produces results comparable with those of subjective testing. Objective measurements can be performed in an intrusive or non-intrusive way. Intrusive measurements require access to the source then compares the original and impaired videos. Full reference and reduced reference video quality measurements are both intrusive [4]. Quality metrics such as peak signal-to-noise ratio (PSNR), structural similarity index measurement [5], video quality metric (VQM) [6] and perceptual evaluation of video quality (PEVQ_) [7] are full reference metrics. VQM and PEVQ_ are commercially used and are not publicly available. Non-intrusive methods (reference-free), on the other hand do not require access to the source video. Non-intrusive methods are either signal or parameter based. Non-intrusive methods are preferred to intrusive analysis as they are more suitable for online quality prediction/control.

Recently, there has been work on video quality prediction. The authors in [8–10] predicted video quality for mobile/wireless networks taking into account the application level parameters only, whereas the authors in [11] used the network statistics to predict video quality. Also, the authors in [12] have proposed a video quality measurement metric (rPSNR) developed from network packet loss conditions. In [13], the authors have proposed a model to measure temporal artefacts on perceived video quality in mobile video broadcasting services. The authors in [14] have proposed a reference-free quality index, which uses an effective human visual system model. Their proposed index is able to assess several spatio-temporal distortions. We proposed in [15] content-based video quality prediction models over WLANs that combined both the application and network-level parameters. It is apparent that video transmission over UMTS network may be subject to QoE degradation because of bandwidth limitation when supporting large number of users. This has been addressed by a large number of researchers in [16–19]. In [20], the authors show that UMTS radio link of acknowledged mode outperforms the unacknowledged mode for video transmission. Similarly, in [21–23] the authors outline the transmission requirements and the performance of UMTS-dedicated channels of UMTS networks for video streaming. In [24], the authors have proposed a mechanism for congestion control for video transmission over UMTS networks, whereas in [25] an error detection scheme is proposed for H.264 encoded videos. In [26] transcoding is used to adapt video content transmitted over UMTS networks. Most of the current work is limited to improving the radio channel. However, very little work has been done on the impact of different types of content on end-to-end video quality over UMTS networks.

There are many parameters that affect video quality and their combined effect is unclear, but their relationships are thought to be non-linear. Artificial neural networks (ANNs) can be used to learn this non-linear relationship which mimics human perception of video quality. ANNs have been widely used in assessing the video quality. The authors in [27, 28] have developed neural-network models to predict video quality based on application and network parameters. They did not consider the different video contents types in developing the neural network models and their work was only limited in fixed Internet protocol (IP) networks. A recent work has also shown the importance of video content in predicting video quality. In [29] features representing video content were used to predict video quality together with other application-level parameters such as SBR and FR. However, this work did not consider any network-level parameters in video quality prediction. In [30], we proposed an adaptive neural fuzzy inference system (ANFIS)-based prediction model that considers both application and network-level parameters over WLAN. Most of the work listed is over IP networks and WLANs. However, video quality prediction over UMTS networks is still less researched and hence, the motivation of our work.

There is a need for an efficient, non-intrusive video quality prediction model for technical and commercial reasons over UMTS networks. The model should predict perceptual video quality to account for interactivity. In this paper, we predict the video quality through a reference-free parameter-based learning model. The work presented here is an extension of previous work [30] on video quality prediction over IEEE 802.11b WLAN standards to propose one model for all contents (as compared to three) and extend to third-generation UMTS networks.

In this paper, first, through statistical analysis we study the impact of QoS parameters both in the application and physical layers on end-to-end perceived quality for all contents over UMTS networks. In particular, we address the problem of evaluating visual quality of low SBR videos under different network conditions (captured by introducing BLER and link bandwidth (LBW)). Based on our MOS results, we perform thorough statistical analysis to study the impact of the five QoS parameters and point out some interesting observations. We believe that our observations add to the existing findings in video quality assessment, and thus have applications in video adaptation for scalable video over mobile networks. Secondly, we develop an ANFIS-based hybrid ANN learning model for perceived video quality prediction as it combines the advantages of neural networks and fuzzy systems. We use

ANFIS [31] to train the neural network using three distinct content types (CTs) [30] to predict the video quality based on a set of objective parameters. The ANN is validated with three different contents in the corresponding categories. We predict video quality in terms of MOS from both physical and application layer parameters for MPEG4 [32] video streaming over WLAN and H.264 [33] video streaming over UMTS networks. We used CT, FR and SBR as application layer parameters and BLER, (packet error rate (PER) for WLAN) and LBW as physical layer parameters. The proposed test bed is based on simulated network scenarios using a network simulator (NS2) [34] with an integrated tool Evalvid [35] and enhanced UMTS radio access network extension (EURANE) [36]. It gives a lot of flexibility for evaluating different topologies and parameter settings used in this study.

The paper is organised as follows. In Section 2, the video quality assessment problem is formulated. Section 3 presents the ANFIS-based ANN learning model along with the training methods. The testing pre-requisites are outlined in Section 4, whereas Section 5 gives the evaluation set-up and the test sequences for both WLAN and UMTS networks. Section 6 describes the impact of QoS parameters on end-to-end video quality. In Section 7, the performance of the ANFIS-based ANN learning models is evaluated. Section 8 concludes the paper and highlights areas of future work.

2 Problem statement

In multimedia streaming services, there are several parameters that affect the visual quality as perceived by the end users of the multimedia content. These QoS parameters can be grouped under application layer QoS and physical layer QoS parameters. Therefore in the application layer perceptual QoS of the video bitstream can be characterised as

$$\text{Perceptual QoS} = f(\text{CT, SBR, FR, codec type, resolution}, \dots)$$

whereas in the physical layer it is given by

$$\text{Perceptual QoS} = f(\text{BLER, LBW, delay, latency, jitter}, \dots)$$

It should be noted that the encoder and content dimensions are highly conceptual. In this research, we chose H.264 as the encoder type for UMTS and MPEG4 for WLAN. H.264 was chosen for UMTS as it is the recommended codec for low bitrates. We used our previously defined classification function [30] to classify the video contents based on their spatial and temporal features. In the application layer, we chose SBR, FR and CT and in the physical layer we chose BLER and LBW as QoS parameters. A single MOS value is used to describe the perceptual quality. Therefore MOS in the application layer is given as MOS^A, whereas MOS in the physical layer is

given by MOS^P as

$$\text{MOS}^A = \{\text{CT, SBR, FR}\} \quad \text{and} \quad \text{MOS}^P = \{\text{BLER, LBW}\}$$

Assuming MOS^A and MOS^P have the same scale then the overall MOS is a function of

$$\text{MOS} = f\{\text{MOS}^A, \text{MOS}^P\}$$

Previous research findings substantiate that the video quality is affected by a number of parameters and their relationships are thought to be non-linear. Therefore based on non-linear regression modelling we have shown in [37] that

$$\text{MOS} \propto (f(\text{MOS})^A) \quad \text{and} \quad \text{MOS} \propto \frac{1}{f(\text{MOS}^P)}$$

Therefore the overall MOS is then given by

$$\text{MOS} = k + \frac{f(\text{MOS}^A)}{(f(\text{MOS})^P)}$$

where k is a constant, $(f(\text{MOS})^A)$ is measured in terms of SBR, FR and CT and $(f(\text{MOS})^P)$ is measured in terms of BLER and mean burst length for two-state Markov model.

In this paper, we evaluated the impact of QoS parameters both in the application and physical layers and confirmed the choice of parameters in the development of the learning models. Therefore the main contributions of the paper are two-fold.

• Evaluation of the impact of QoS parameters on end-to-end video quality for H.264 encoded video.

• Development of new and efficient learning models to predict video quality non-intrusively avoiding time-consuming subjective tests.

3 ANFIS-based ANN learning models

3.1 Background to ANFIS-based ANN

ANFIS uses a hybrid learning procedure and can construct an input−output mapping based on both human knowledge (in the form of fuzzy if−then rules) and stipulated input−output data pairs. A two input ANFIS architecture as shown in Fig. 1 is an adaptive multilayer feedforward network in which each node performs a particular function on incoming signals as well as a set of parameters pertaining to this node.

The entire system architecture in Fig. 1 consists of five layers, namely, a fuzzy layer, a product layer, a normalised layer, a defuzzy layer and a total output layer. The two inputs are x and y. The output is f. For a first-order
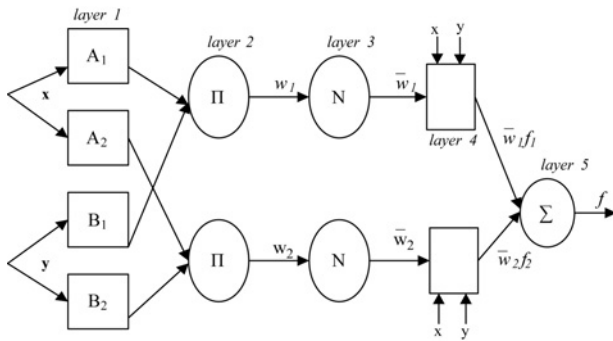
Figure 1 *ANFIS architecture [31]*

Sugeno fuzzy model, a typical rule set with two fuzzy if–then rules can be expressed as:

- *Rule 1:* If x is A_1 and (y is B_1) then $f_1 = p_1 x + q_1 y + r_1$

- *Rule 2:* If x is A_2 and (y is B_2) then $f_2 = p_2 x + q_2 y + r_2$ where p_1, p_2, q_1, q_2, r_1 and r_2 are linear parameters, and A_1, A_2, B_1 and B_2 are non-linear parameters.

$$f = \frac{w_1 f_1 + w_2 f_2}{w_1 + w_2}$$

From Fig. 1, Layer 1 is the fuzzy layer, in which x and y are the input of nodes. A_1, A_2, B_1 and B_2 are the linguistic labels used in the fuzzy theory for dividing the membership functions (MFs). The membership relationship between the output and input functions of this layer can be expressed as

$$O_{1,i} = \mu A_i(x), \quad i = 1, 2$$
$$O_{1,j} = \mu B_j(y), \quad j = 1, 2 \tag{1}$$

where $O_{1,i}$ and $O_{1,j}$ denote the output functions and μA_i and μB_j denote the MFs. Layer 2 is the product layer that consists of two nodes labelled Π. The outputs w_1 and w_2 are the weight functions of the next layer. The output of this layer is the product of the input signal, which is defined as follows

$$O_{2,i} = w_i = \mu A_i(x) \times \mu B_i(y) \quad i = 1, 2 \tag{2}$$

where $O_{2,i}$ denotes the output of Layer 2.

The third layer is the normalised layer, whose nodes are labelled N. Its function is to normalise the weight function in the following process

$$O_{3,i} = \bar{w}_i = \frac{w_i}{w_1 + w_2} \quad i = 1, 2 \tag{3}$$

where $O_{3,i}$ denotes the Layer 3 output.

The fourth layer is the defuzzy layer, whose nodes are adaptive. The output equation is $\bar{w}_i(p_i x + q_i y + r_i)$, where p_i, q_i and r_i denote the linear parameters or so-called

consequent parameters of the node. The defuzzy relationship between the input and output of this layer can be defined as

$$O_{4,i} = \bar{w}_i f_i = \bar{w}_i(p_i x + q_i y + r_i) \tag{4}$$

where $O_{4,i}$ denotes the Layer 4 output. The fifth layer is the total output layer, whose node is labelled $\sum$. The output of this layer is the total of the input signals, which represents the shift decision result. The results can be written as

$$O_{5,i} = \text{overall output} = \sum_i \bar{w}_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \tag{5}$$

where $O_{5,i}$ denotes the Layer 5 output [29].

3.2 Introduction to the models

The aim is to develop two ANFIS-based learning models to predict perceived video quality for three distinct CTs from both physical and application layer parameters for MPEG4 and H.264 video streaming over WLAN and UMTS access networks. The functional block of the proposed model is shown in Fig. 2.

For the tests, we selected three different video sequences representing slow moving content to fast moving content as classified in our previous work [30]. The video sequences were of quarter common intermediate format (QCIF) resolution (176×144) and encoded in H.264 format with an open source JM software [33] encoder/decoder for UMTS access network and in MPEG4 [32] for WLAN access network. The three video clips were transmitted over simulated UMTS and WLAN access networks using NS2 simulator. The application layer parameters considered are CT, FR and SBR. The physical layer parameters are BLER and LBW for UMTS and PER and LBW for WLAN.

3.3 ANFIS architecture

The corresponding equivalent ANFIS architecture for the two learning models developed over WLAN and UMTS access networks is shown in Fig. 3. There are five inputs as CT, FR, SBR, LBW and BLER/PER and the output is the MOS value. The MFs for the ANFIS learning model
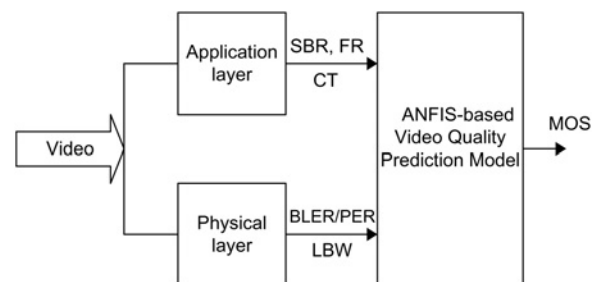

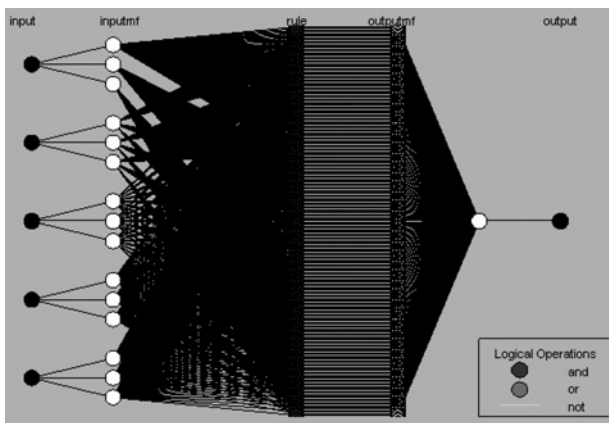**Figure 2** *Functional block of proposed ANFIS-based model*

Figure 3 *ANFIS architecture of the proposed model*

given in Fig. 3 as *inputmf* are given in Figs. 4*a* and *b* for UMTS and WLAN.

The number of MF is chosen to be three for all five inputs and their operating range depends on the five inputs. For example, the operating range for the input of SBR is from [0, 105], whereas the CT ranging from [1, 3] represents the three different types of content. Note that throughout the simulation all the MFs used are generalised bell function defined in [31] as

$$\mu_a = \frac{1}{1 + [([x - c/a])^2]^b}$$

which contains the fitting parameters *a*, *b* and *c*. Each of these parameters has a physical meaning: *c* determines the centre of the corresponding MF; *a* is the half-width and *b* (together

with *a*) controls the slopes at the crossover points (where MF value is 0.5) in Figs. 4*a* and *b*.

3.4 Training and validating of ANFIS-based models

For ANNs, it is not a challenge to predict patterns existing on a sequence with which they were trained. The real challenge is to predict sequences that the network did not use for training. However, the part of the video sequence to be used for training should be 'rich enough' to equip the network with enough power to extrapolate patterns that may exist in other sequences. Three different CTs representing different scenarios from slow movement to fast moving sports clips [30] are chosen for training purposes and three different video clips for validation purposes. The ANFIS-based ANN model were trained with the three distinct CTs of 'Akiyo', 'Foreman' and 'Stefan' (see Tables 2 and 3) and validated by three different CTs of 'Suzie', 'Carphone' and Football' in the corresponding content categories. Snapshots of the six video clips used for training and validation are given in Fig. 5.

The data selected for validation were one-third that of testing with different parameter values to that given in Tables 2 and 3. In total there were around 450 encoded test sequences for training and 150 encoded test sequences for validation for each model.

4 Testing pre-requisites

In this section, we describe the testing pre-requisites for WLAN and UMTS for multimedia streaming.
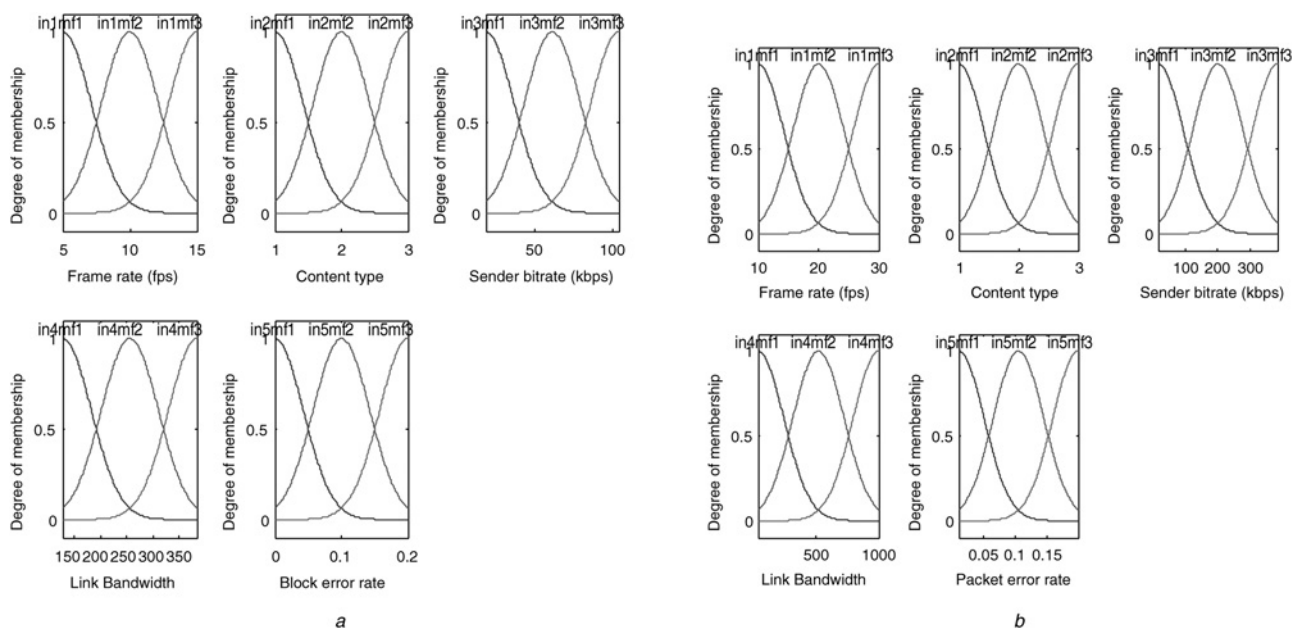

Figure 4 *MFs of the five inputs of the proposed model*
a MFs for five inputs over UMTS
b MFs for five inputs over WLAN

Figure 5 *Snapshots of six CTs for training and validation of the model*

4.1 IEEE 802.11 WLAN

The IEEE 802.11 set of specifications are wireless standards that specify an over-the-air interface between a wireless client and a base station (access point) as well as among wireless clients. Two basic operating modes are defined as infrastructure mode and *ad hoc* mode. The infrastructure mode allows clients to roam between access points, while roaming across routers is prohibited. The *ad hoc* mode allows individual nodes to participate in a peer-to-peer communication without an access point. In this paper, we have considered only the infrastructure mode.

The standard has currently three variations widely deployed. The 802.11b operates at the 2.4 GHz band and has a maximum theoretical data rate of 11 Mbps, but operates also on 1, 2 and 5 Mbps. The 802.11a and g operate at the 5 and 2.4 GHz bands, respectively, and both have a theoretical data rate of 54 Mbps. Using different modulation schemes they can also operate on the low scales of 6, 10, 12, 18, 36 and 48 Mbps. In this paper, our simulations are based on IEEE 802.11b with 11 Mbps.

4.2 UMTS network and architecture

3G (third generation) systems are intended to provide a global mobility with wide range of services including telephony, paging, messaging, Internet and broadband data. UMTS offers teleservices and bearer services, which provide the capability for information transfer between access points. It is possible to negotiate the characteristics of a bearer service at session or connection establishment and renegotiate them during the session or connection. Bearer services have different QoS parameters for maximum transfer delay, delay variation and bit error rate. UMTS network services have different QoS classes: conversational class (voice, video telephony, video gaming), streaming class (multimedia, video on-demand (VOD), webcast), interactive class (web browsing, network gaming, database access) and background class (e-mail, SMS, downloading).

The offered data rate targets are 144 Kbps for satellite and rural outdoor, 384 Kbps for urban outdoor and 2048 Kbps for indoor and low-range outdoor. These are the maximum theoretical values in each environment for downlink speeds. The actual data rates may vary from 32 Kbps, for a single voice channel, to 768 Kbps in urban low-speed connections depending always on the class of service supported.

Fig. 6 shows a simplified architecture of UMTS for packet-switched operation [1], which consists of one or several user equipments (UEs), the UMTS terrestrial radio access network (UTRAN) and the core network. The UTRAN is composed of node Bs connected to a radio network controller (RNC). The core network, which is the backbone of UMTS, comprises the serving general packet radio service (GPRS) support node (SGSN) and the gateway GPRS support node (GGSN). The SGSNs route packets to and from UTRAN, while GGSNs interface with external IP networks. UE, which is a mobile station, is connected to node B over the UMTS radio interface.

Fig. 6 shows the example of H.264 video transmission from a fixed Internet node to the mobile user. The packets leave the RNC and arrive at the node B they queue up in order to be broken down into smaller size packets. Every packet data convergence protocol (PDCP) packet data unit (PDU) are segmented into multiple radio link controller (RLC) PDUs of fixed size. Each of these PDUs fit into a block in order to be transmitted over the air. The RLC model supports both acknowledged modes (AMs) and unacknowledged modes (UMs) of operation. The AMs guarantees delivery by retransmitting erroneous RLC blocks at the expense of transfer delay, whereas the UM provides unreliable but timely delivery of RLC blocks. For this reason, in this paper we have considered the AM only.

4.3 Multimedia streaming traffic

Multimedia applications are continuously growing in popularity. Real-time multimedia traffic consists of one or more media streams and can be characterised by strict delay requirements while can tolerate some losses.

Video streams compressed with encoders like MPEG4 or H.26x exhibit large variations in their data rates, something which makes their management in a packet-based best effort network like IP extremely difficult. It is crucial to
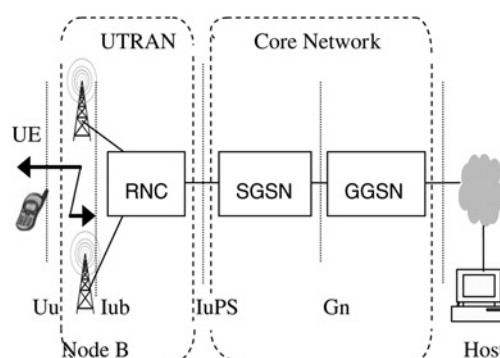

Figure 6 *UMTS reference architecture*

predict the QoS degradation that may be experienced by multimedia applications over wireless access networks. Video transmissions applications need to be responsive to dynamic changes and different demands [24]. In the presence of packet loss, video quality becomes highly time-variant [24, 25]. One of the significant problems that video streaming face is the unpredictable nature of the Internet in terms of the SBR, end-to-end delay and loss variation.

5 Simulation set-up

This section describes the simulation set-up, test sequences and variable test parameters for IEEE 802.11b WLAN and UMTS networks.

5.1 Simulation set-up for IEEE 802.11b WLAN

For the tests, we selected three different video sequences of QCIF resolution (176×144) and encoded in MPEG4 format with an open source ffmpeg [30] encoder/decoder with a group of pictures (GOP) pattern of IBBPBBPBB. Each GOP encodes three types of frames – intra (I) frames are encoded independently of any other type of frames, predicted (P) frames are encoded using predictions from preceding I or P frames and bi-directional (B) frames are encoded using predictions from the preceding and succeeding I or P frames.

The experimental set-up is given in Fig. 7. There are two sender nodes, namely constant bitrate (CBR) background traffic and MPEG4 video source. Both the links pass traffic at 10 Mbps over the Internet which in turn passes the traffic to another router over a variable link. The second router is connected to a wireless access point at 10 Mbps and further transmits this traffic to a mobile node at a transmission rate of 11 Mbps 802.11b WLAN. The delay is fixed at 1 ms. No packet loss occurs in the wired segment of the video delivered path. The maximum transmission packet size is 1024 bytes. The video packets are delivered with the random uniform error model. The CBR rate is fixed to 1 Mbps to simulate realistic scenario. The PER is set in the range of $0.01-0.2$ with 0.05 intervals. To account for different packet loss patterns, ten different initial seeds for random number generation were

Table 1 PSNR to MOS conversion

PSNR, dB	MOS
>37	5
31–36.9	4
25–30.9	3
20–24.9	2
<19.9	1

chosen for each PER. All the results generated in this paper were obtained by averaging over these ten runs.

All the experiments in this paper were conducted with an open source framework Evalvid [35] and network simulator tool NS2 [34]. Video quality is measured by taking the average PSNR over all the decoded frames. MOSs are calculated based on the PSNR to MOS conversion from Evalvid [35]. The mapping is given in Table 1.

The motivation for using an objective method as opposed to subjective testing was because subjective testing is time-consuming and expensive. NS2 integrated with Evalvid simulation platform gave a lot of flexibility in choice of parameters. However, as PSNR is not a good reflector of visual quality, we have conducted controlled subjective tests over UMTS network to validate the model. For details see Section 7.3.

5.2 Test sequences and variable test parameters for IEEE 802.11b WLAN simulation

For quality evaluation, we used a combination of application and network-level parameters as FR, SBR, LBW and PER. The video sequences along with the combination parameters chosen are given in Table 2.

In the application level, we considered the following: (i) The FR – the number of frames per second. It takes one of the three values as 10, 15 and 30 fps; (ii) The SBR – the rate of the encoders output. It is chosen to take 18, 44, 80, 104 and 512 kbps.
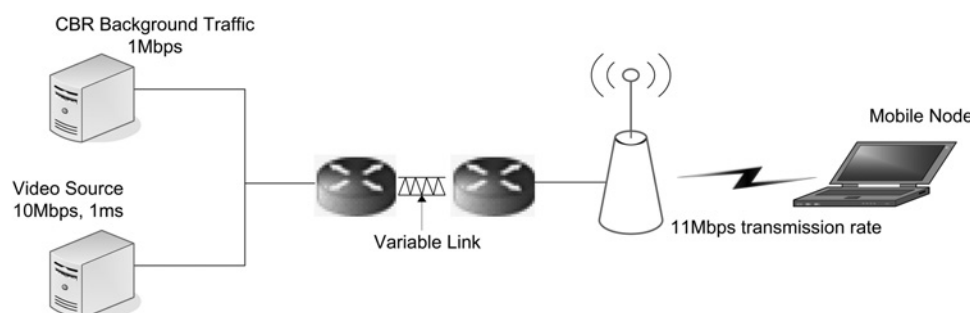

Figure 7 Simulation set-up

Table 2 Simulation parameters for WLAN

Video sequences	FR, fps	SBR, kbps	Link BW, kbps	PER
Akiyo, Foreman	10, 15, 30	18, 44, 80	32, 64, 128, 256, 384	
Suzie, Carphone	15, 30	44, 80	128, 256, 384	0.01, 0.05, 0.1, 0.15, 0.2
Stefan	10, 15, 30	80, 104, 512	256, 384, 512, 768, 1000	
Football	15, 30	104, 512	512, 768, 1000	

In the network level, we considered the following: (i) The LBW: the variable bandwidth link between the routers (Fig. 7). It takes the values of 32, 64, 128, 256, 384, 512, 768 and 1000 kbps. (ii) PER: the simulator (NS2) [34] drops packet at regular intervals using the random uniform error model, taking five values as 0.01, 0.05, 0.1, 0.15 and 0.2. It is widely accepted that a loss rate higher than 0.2 (20%) will drastically reduce the video quality.

5.3 Simulation set-up for UMTS network

The network topology is modelled in the UMTS extension for the NS2 [34], namely enhanced UMTS radio access network extension (EURANE) [36] integrated with Evalvid [35] for H.264 video streaming. H.264 codec is chosen as opposed to MPEG4 as it the recommended codec for low bitrate transmission.

The simulation model is given in Fig. 8. It consists of a streaming client and a server. In the simulation, the UE is a streaming client and a fixed host is the streaming server located in the Internet. The addressed scenario comprises a UMTS radio cell covered by a node B connected to an RNC. The simulation model consists of a UE connected to downlink dedicated physical channel (DPCH).

As the main aim of the simulation was to investigate the impact of the radio interface (UMTS network) on the quality of streaming H.264 video with varying SBR, no packet losses occur either on the Internet or on the UMTS core network. In Fig. 8 the links between the two nodes are labelled with their bitrate (in bits per second) and delay (in seconds). Each link capacity was chosen so that the radio channel is the connection bottleneck. Consequently,

the functionality of SGSN and GGSN was abstracted out and modelled as traditional NS nodes since they are wired nodes and in many ways mimic the behaviour of IP router. Currently no header compression technique is supported in the PDCP layer.

From the 3GPP recommendations, we find that for video streaming services, such as VOD or unicast Internet protocol television services, a client should support H.264 (advanced video coding) baseline profile up to the level 1.2. [33]. As the transmission of video was for mobile handsets, all the video sequences are encoded with a QCIF resolution. The considered frame structure is IPP for all the sequences, since the extensive use of I frames could saturate the available data channel. From these considerations, we set up the encoding features as shown in Table 3.

5.4 Test sequences and variable test parameters for UMTS simulation

UMTS physical model is responsible for transmitting blocks over the physical channels. The channel bitrates and the transmission time interval associated with the channel considered in the simulation are shown in Table 2. The UMTS downlink bitrate has one of the three values as 128, 256 and 384 kbps. Since the physical layer passes the block to the medium access layer together with the error indication from the cyclic redundancy check, the output of the physical layer can be characterised by the overall probability of the block error – referred to as BLER in this paper. Therefore an error model based on the uniform distribution of block errors was used in the simulation. The BLER ranges between 0 and 20% in our simulation. FR
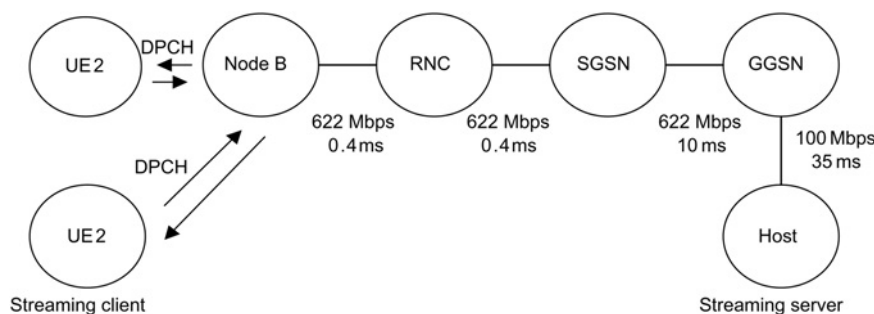

Figure 8 UMTS network topology

Table 3 Simulation parameters for UMTS

Input parameter	Value
UMTS physical channel type	DPCH
downlink bitrate, kbps	128, 256, 384
uplink bitrate, kbps	64
downlink TT1, ms	10
uplink TT1, ms	20
packet size (bytes)	1000
UDP header size (bytes)	8
IP header size (bytes)	20
RLC mode	acknowledged
encoder Profile/Level IDC	(66, 11), baseline profile, level 1.1
sequence type	IPPP
entropy coding method	CAVLC
BLER	0−20%
error model	uniform distribution
SBR	18, 44, 80, 104 kbps
FR	5, 10, 15 fps

ranges from 5 to 15 fps, whereas the SBR ranges from 18 to 104 kbps for all contents. See Table 3 for details.

6 Impact of QoS parameters on video quality over UMTS network

In order to thoroughly study the influence of different QoS parameters on MOS, we perform analysis of variance (ANOVA) [38] on the MOS data set. Table 4 shows the results of the ANOVA analysis.

We performed five-way ANOVA to determine if the means in the MOS data set given by the five QoS parameters differ when grouped by multiple factors (i.e. the impact of all five parameters on MOS). Table 4 shows the

results, where the first column is the sum of squares, second column is the degrees of freedom associated with the model, the third column is the mean squares, that is, the ratio of sum of squares to degrees of freedom. The fourth column shows the F-statistic and the fifth column gives the p-value, which is derived from the cumulative distribution function of F [38]. A small p-value ($p \leq 0.01$) indicates that the MOS is substantially affected by a variation of the corresponding parameter. Furthermore, based on the magnitudes of p-values, we can make a further claim that CT and LBW (p-value = 0) impacts the MOS results the most, followed by BLER and then SBR, while FR has the least influence. As the MOS is found to be mostly affected by CT and LBW, we further categorise the CT and LBW using the multiple comparison test based on Tukey−Kramer's honestly significant difference criterion [39]. The results of comparison test for CT and LBW are shown in Figs. 9a and b, where the centre and span of each horizontal bar indicate the mean and the 95% confidence interval, respectively.

Our studies numerically substantiate the following observations reported in previous studies of video quality assessment:

• The most important QoS parameter in the application layer is the CT. Therefore an accurate video quality prediction model must consider all CTs. The values of FR and SBR are independent from the CT and therefore cannot provide accurate estimation of quality.

• The optimal combination of SBR and FR that gives the best quality is very much content dependent and varies from sequence to sequence. We found that for slow moving content FR = 5 and SBR = 18 kbps gave acceptable quality, however as the spatio-temporal activity of the content increased this combination gave unacceptable quality under no network impairment. This is shown by Figs. 10a and b. From Fig. 10a when CT = 3 (Stefan), SBR = 20 kbps for an FR of 10 fps MOS < 1. Similarly, from Fig. 10b when FR = 5 fps for CT of Stefan at an SBR of 48 kbps MOS < 1. However, when CT = 2/1 (Foreman/Akiyo) then for the same conditions MOS increases to 2.5. This clearly shows that as the spatio-temporal activity of the content increases low

Table 4 Five-way ANOVA on MOS

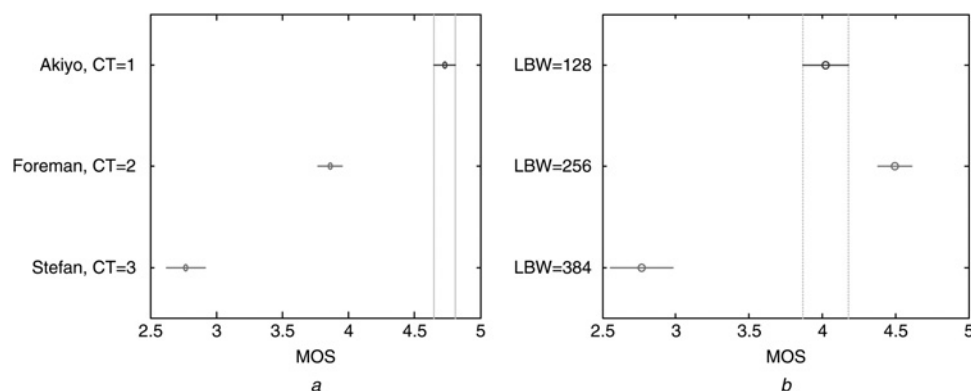| Parameters	Sum of squares	Degrees of freedom	Mean squares	F-statistic	p-value
CT	15.052	2	15.052	362.71	0
FR	0.2598	2	0.1299	3.13	0.0489
SBR	4.462	2	2.23098	3.99	0.0217
BLER	0.9926	4	0.2481	5.98	0.0003
LBW	35.381	2	17.6905	79.43	0

Figure 9 *Multiple comparison test results*

a Multiple comparison test for CT against MOS
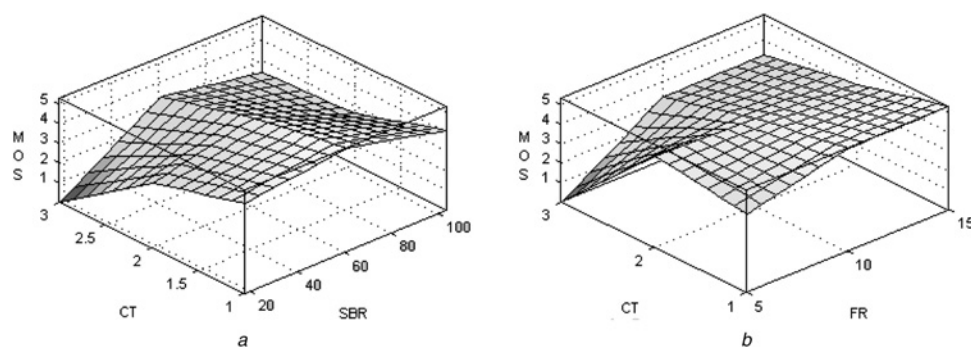*b* Multiple comparison test for LBW against MOS

Figure 10 *3D plots of MOS vs SBR/FR for all contents*

a MOS against CT and against SBR
b MOS against CT and against FR

FRs and SBRs give very low quality. However, with slow-to-medium spatio-temporal activity, the low FR–SBR combination gives acceptable quality. Hence the choice of SBR and FR is very much dependent on the type of content. Also knowing the initial encoding SBR saves useful bandwidth resources.

• The ANOVA results showed that the QoS parameter of LBW had a significant impact on quality. In real systems, the impact of LBW is generally measured by BLER. Therefore an accurate video quality prediction model must take into account the influence of physical layer in addition to application layer parameters.

• The impact of physical layer parameters of LBW and BLER vary depending on the type of content. For slow moving content BLER of 10% gives acceptable quality, however, for fast moving content for the same BLER the quality is completely unacceptable. Therefore the impact of physical layer QoS parameters is very much content dependent as well. This is explained in Figs. 11*a* and *b*. The CT is defined in the range of [1, 3] from slow moving to fast moving sports type of content. From Fig. 11*a*, we observe that as the activity of the content increases the impact of BLER is much higher. For example, for 20% BLER, CT of slow to medium type gives very good MOS,

whereas as the content activity increases, MOS reduces to 3. From Fig. 11*b* we observe that if the LBW is 128 kbps then quality is low because of network congestion (for content encoded at SBR close to 128 kbps). The impact of LBW is normally measured by BLER in real systems.

7 Evaluation of the proposed ANFIS-based learning models

The aim was to develop ANFIS-based learning models to predict video quality. We trained the ANFIS-based learning model using three distinct CTs and validated them with three different video test sequences in the corresponding content categories representing from low to high spatio-temporal features. The models are trained with objective data (PSNR to MOS) conversion. The models over UMTS network are further validated with subjective results. The accuracy of the proposed models can be determined by the correlation coefficient and the root mean squared error (RMSE) of the validation results.

7.1 Model over IEEE 802.11b WLAN

Fig. 12*a* shows the predicted MOS results for our proposed model over WLAN, whereas Fig. 12*b* shows the validation
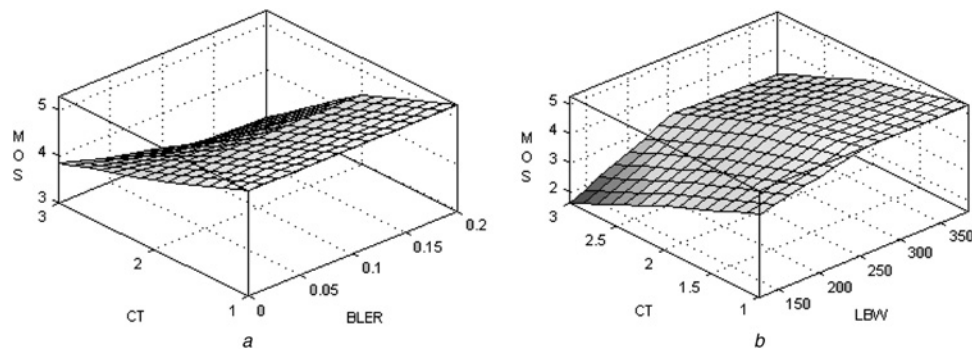
Figure 11 *3D plots of MOS vs BLER/LBW for all contents*
a MOS against CT and against BLER
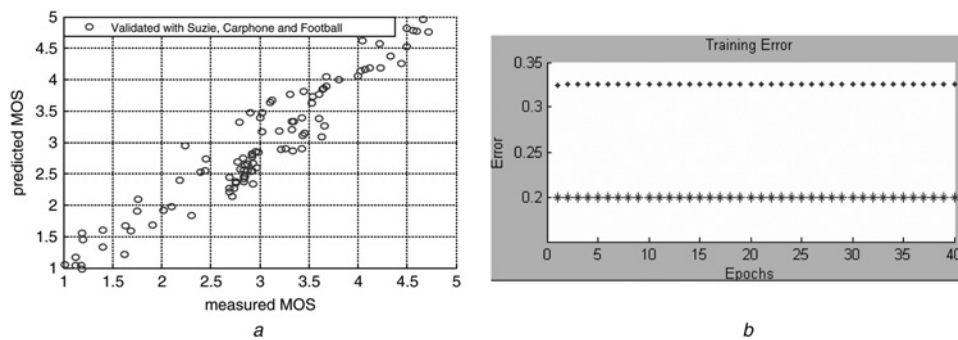*b* MOS against CT and against LBW

Figure 12 *MOS predicted over WLAN*
a Predicted MOS against measured MOS over WLAN
b Validation error against training error over WLAN

error against the training error. The validation results of the proposed ANFIS-based model in terms of the correlation factor R^2 is 90.42% and the RMSE between the predicted and measured MOS for all CTs is 0.31.

7.2 Model over UMTS network

Fig. 13a shows predicted MOS values for our proposed model over UMTS network. In Fig. 13b the validation error is given against the training error. The validation

results of the proposed ANFIS-based model in terms of the correlation factor R^2 is 86.91% and RMSE between the predicted and measured MOS for all CTs is 0.3247.

Fig. 13a shows a number of points together with MOS ~ 5. This is because the error correlation properties of the link layer in UMTS as opposed to WLAN do not have an impact on the quality of the streamed video as long as the IP packet error probability remains unchanged. Therefore for slower moving content
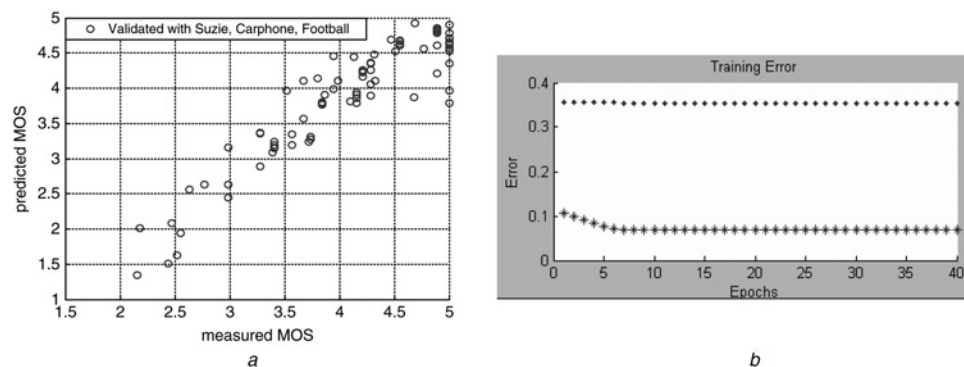

Figure 13 *MOS predicted over UMTS*
a Predicted MOS against measured MOS over UMTS
b Validation error against training error over UMTS

BLER of 20% did not result in quality degradation as compared to faster moving content. PSNR to MOS conversion resulted in a number of points having a MOS of 5 for high PSNR values (>37 dB). However, this is corrected in the subjective MOS values. See Fig. 15 in Section 7.3.

7.3 Validation of UMTS model via subjective testing

In order to validate our proposed model over UMTS networks we conducted subjective tests for the test conditions given in Table 3 for model validation only. The BLER values chosen were 0.01 and 0.2 and the FR was fixed at 10 fps due to time and resource constraints. The assessment method, participant information, testing conditions and the test results are outlined in this sub-section.

7.3.1 Assessment method:
The subjective quality assessment experiment follows ITU-T recommendations [40] and was conducted using the single-stimulus absolute category rating method with a five-point quality scale [40]. The degraded video clips are viewed one at a time and rated independently on a discrete five-level scale from 'bad' (1) to 'excellent' (5). The snapshot of the rating description on subjective test webpage is shown in Fig. 14. The ratings for each test clip are then averaged over all subjects to obtain an MOS. The degraded video clips were randomised; for example, one with BLER of 20% was presented before the sequence of BLER with 1% loss. The voting period was not time-limited. After choosing their

quality rating, assessors had to confirm their choice using the 'submit' button as shown in [41]. This approach gave subjects the possibility to change their mind before committing to their final vote.

7.3.2 Participants:
A total of 16 naïve viewers participated in the experiment, seven males and nine females. This conforms to the minimum number of viewers specified by ITU-T recommendations [40]. Participants were recruited from within the university. The chosen group of test persons ranged different ages (4 over 35, 2 between 26 and 30 and 10 between 18 and 25). They did not have expertise in video processing and quality assessment and had a non-technical background. None of these assessors had participated in a subjective quality assessment experiment. They were advised to go to the set URL where the tests were located and read the instructions to carry out the test. The URL is given in [41]. All subjects assessed all degraded video sequences in the test.

7.3.3 Testing environment:
The laboratory had calibrated 20-inch computer LCD monitor (Philips 200WB7) to display the video sequences. The display had a native resolution of 1280 × 1024 pixels and colour quality selected as highest (32 bit). The room had a white background. Participants provided their ratings electronically using the computer mouse. A specially designed webpage based on asp.net [41] was created that contained the degraded video sequences.

7.3.4 Test results:
Fig. 15a shows predicted MOS values for our proposed model over UMTS network against the

Rating	Definition	Description
5	Excellent	e.g. a perfect video clip
4	Good	e.g. Video clip with very good quality
3	Fair	e.g. video clip with acceptable quality i.e. some loss of quality but overall image acceptable
2	Poor	e.g. low quality, image distorted, hard to understand the video
1	Bad	communications breakdown

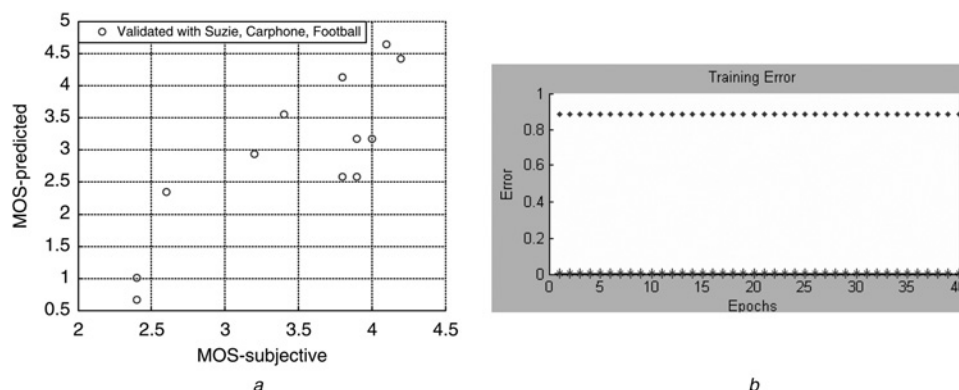**Figure 14** *Snapshot of subjective rating description to participants*

Figure 15 *MOS predicted over UMTS for subjective results*
a Subjective MOS against predicted MOS
b Validation error against training error

subjective MOS values. In Fig. 15b the validation error is given against the training error. The validation result of the proposed ANFIS-based model in terms of the correlation factor R^2 is 82.9% and the RMSE between the predicted and subjective MOS for all CTs is 0.8896. The training error is quite low compared to the validation error. This is due to a number of reasons. The validation conditions are only 12, which are very low compared to the training conditions. Also, the training of the model is done via objective MOS obtained from PSNR conversion, and the validation of model is done via subjective MOS. For an accurate reflection of the model, in future we will run extensive subjective tests that will allow us both to train and validate the model subjectively. However, a controlled subjective experiment conducted in this manner proves the concept.

7.4 Comparison of the three models

The models proposed in this paper are reference-free. The comparison of the three models in terms of the correlation coefficient (R^2) and RMSE is given in Table 5. The ANFIS-based WLAN model outperformed the model over UMTS in terms of prediction accuracy for objective MOS data. However, the difference in performance was not massive. The performance of the ANFIS-based model over UMTS was slightly worse because the error correlation properties of the link layer do not have an impact on the quality of the streamed video as long as the IP packet error probability remains unchanged. The UMTS model validation with subjective data performs worse than the objective MOS data. However, because of time and resource constraints the model is only validated with subjective test data. It is trained with objective MOS. Moreover, the test conditions were a lot less compared to those from objective MOS. Nevertheless, the results give an indication that the model performs well in terms of correlation coefficient. The RMSE is high mainly because the validation conditions were only 12. The subjective tests were performed to prove concept. However, in future extensive subjective tests will be carried out both for model training and validation. Also, in future we will consider the burstiness of the network by simulating two-state Markov model compared to the random uniform model that we simulated in this paper.

Compared to our previous work [30], where we proposed three models for the three CTs, both models performs very well (∼90% correlation). We feel that the choice of parameters is crucial in achieving good prediction accuracy. Parameter such as LBW in real systems is measured in

terms of block error/packet loss and delay. However, in a simulation system it was interesting to capture the impact of LBW. Also, in the application layer the SBR has a bigger impact on quality than FR, whereas if FR is reduced too low, for example, 5 fps then FR has a bigger impact on quality. Finally, to predict video quality CT is the most important QoS parameter. We found that faster moving content gives low MOS scores over UMTS compared to WLAN. This could be due to the bandwidth restriction over UMTS network for faster moving CTs. Also, contents with less movement require low SBR to that of higher movement to give acceptable quality.

8 Conclusions

In this paper, we presented reference-free learning models based on ANFIS for MPEG4 and H.264 video streaming to predict video quality over WLAN and UMTS networks in terms of MOS, respectively. We obtained good prediction accuracy ($R^2 \sim 90$%) with unseen data set. We further investigated the combined effects of application and physical layer parameters on end-to-end perceived video quality and analysed the behaviour of video quality for wide range of variations for a set of selected parameters over UMTS access networks. We observed that CT in the application layer and LBW in the physical layer are the most important QoS parameters. However, in real systems as LBW is generally measured in terms of BLER and delay, we conclude that video CT is the most important QoS factor as the impact of physical layer parameters is very much content-dependent too. Further, from the ANFIS-based learning models proposed, our results demonstrate that it is possible to predict the video quality if the appropriate parameters are chosen. The subjective data validates our proposed model over UMTS network. Hence, our results confirm that the proposed ANFIS-based learning model is a suitable tool for video quality prediction for the most significant video streaming CTs. This study should help in the development of a reference-free video prediction model and has applications in video adaptation for scalable video over UMTS networks. Bitstream scalability for video is a desirable feature for many multimedia applications. The need for scalability arises from graceful degradation owing to transmission requirements, or adaptation needs. To fulfil these requirements, the thorough understanding of QoS parameters is required. Therefore from our results we can see that a slow moving video encoded at 18 kbps can be sent if transmission errors are high. However, for fast moving contents the quality degradation due to encoding has to be offset against the degradation obtained by transmission errors.

Our future work will focus on extensive subjective testing to train and validate the proposed models. We will apply our results to adapt the video SBR and hence optimise bandwidth for specific CT. Also, they should help in developing effective handover strategies in future next generation networks.

Table 5 Comparison of the proposed ANFIS-based models

Models	R^2, %	RMSE
WLAN	90.42	0.31
UMTS (objective MOS)	86.91	0.3247
UMTS (subjective MOS)	82.9	0.8896

9   Acknowledgment

10   References

[1]   3GPP: Available at http://www.3gpp2.org

[2]   DTR 0001: 'Telecom and Internet converged services and protocols for advanced networking (TISPAN)', Release 1 Definition

[3]   ITU-T Rec P.800: 'Methods for subjective determination of transmission quality', 1996

[4]   Video quality experts group, multimedia group test plan, Draft version 1.8, December 2005, www.vqeg.org

[5]   WANG Z., LU L., BOVIK A.: 'Video quality assessment based on structural distortion measurement', *Signal Process.: Image Commun. Special issue on* 'Objective video quality metrics', 2004, 19, (2), pp. 121–132

[6]   http://www.its.bldrdoc.gov/vqm/

[7]   http://www.pevq.org

[8]   YAMAGISHI K., HAYASHI T.: 'Opinion model using psychological factors for interactive multimodal services', *IEICE Trans. Commun.*, 2006, **E89-B**, (2), pp. 281–288

[9]   KOUMARAS H., KOURTIS A., LIN C., SHIEH C.: 'A theoretical framework for end-to-end video quality prediction of MPEG-based sequences'. Third Int. Conf. on Networking and Services, 19–25 June 2007

[10]   YAMAGISHI K., TOMINAGA T., HAYASHI T., TAKASHI A.: 'Objective quality estimation model for videophone services', *NTT Tech. Rev.*, 2007, **5**, (6)

[11]   CALYAM P., EKICIO E., LEE C., HAFFNER M., HOWES N.: 'A gap-model based framework for online VVoIP QoE measurement', *J. Commun. Netw.*, 2007, **9**, (4), pp. 446–456

[12]   TAO S., APOSTOLOPOULOS J., GUERIN R.: 'Real-time monitoring of video quality in IP networks'. Proc. ACM NOSSDAV, 2005

[13]   HUYNH-THU Q., GHANBARI M.: 'Temporal aspect of perceived quality in mobile video broadcasting', *IEEE Trans. Broadcast.*, 2008, **54**, (3), pp. 641–651

[14]   MASSIDDA F., GIUSTO D., PERRA C.: 'No reference video quality estimation based on human visual system for 2.5/3G devices', *Proc. SPIE – Int. Soc. Opt. Engng.*, 2005, Vol. 5666, (1), pp. 168–179

[15]   KHAN A., SUN L., IFEACHOR E.: 'Content clustering-based video quality prediction model for MPEG4 video streaming over wireless networks'. Proc. IEEE ICC, Dresden, Germany, 14–18 June 2009

[16]   PLIAKAS T., KORMENTZAS G., SKIANIS C.: 'Scalable video streaming traffic delivery in IP/UMTS networking environments', *J. Multimedia*, 2007, **2**, (2), pp. 37–46

[17]   POLITIS I., DAGIUKLAS T., KOTSOPOULOS S.: 'Study of the QoS of video traffic over integrated 3G-WLAN systems'. Mobimedia'06, Alghero, Sardinia, Italy, 18–20 September 2006

[18]   ANGELIS F., HABIB I., DAVIDE F., NAGHSHINEH M.: 'Intelligent content aware services in 3G wireless networks', *IEEE J. Select. Areas Commun.*, 2005, **23**, (2), pp. 221–234

[19]   SALKINTZIS A., SKYRIANOGLOU D., PASSAS N.:: 'Seamless multimedia QoS across UMTS and WLANs'. Vehicular Technology Conf., 30 May–1 June 2005

[20]   LO A., HEIJENK G., NIEMEGEERS I.: 'Performance evaluation of MPEG4 video streaming over UMTS networks using an integrated tool environment'. Int. Symp. on Performance Evaluation of Computer and Telecommunication Systems, SPECTS, Philadelphia, PA, USA, 24–28 July 2005, pp. 676–682

[21]   ALEXIOU A., BOURAS C., IGGLESIS V.: 'Multimedia transmission over third generation cellular network'. Proc. Tenth Int. IFIP Int. Conf., Colmar, France, 25–27 August 2005

[22]   VASSILIOU V., ANTONIOU P., GIANNAKOU I., PITSILLIDES A.: 'Requirements for the transmission of streaming video in mobile wireless networks'. Part I and II: 16th Int. Conf., ICANN, Athens, Greece, 10–14 September 2006

[23]   BRUMME S.: 'Multimedia streaming on mobile phones'. Lecture on Advanced Data Communications, University of Technology, Sydney, 2004

[24]   ALEXIOU A., ANTONELLIS D., BOURAS C.: 'Adaptive and reliable video transmission over UMTS for enhanced performance', *Int. J. Commun. Syst.*, 2007, **20**, (1), pp. 65–81

[25]   NEMETHOVA O., KARNER W., AL-MOGHRABI A., RUPP M.: 'Cross layer detection for H264 video over UMTS'. Proc. Int. Wireless Summit, Aolborg, Denmark, 18–22 December 2005

[26]   DOGAN S., EMINSOY S., SADKA A., KONDOZ A.: 'Video content adaptation using transcoding for enabling UMA over UMTS'. Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS'2004), Lisbon, Portugal, April 2004

[27]   MOHAMED S., RUBINO G.: 'A study of real-time packet video quality using random neural networks', *IEEE Trans. Circuits Syst. Video Technol.*, 2002, **12**, (12), pp. 1071–1083

[28] FRANK P., INCER J.: 'A neural network based test bed for evaluating the quality of video streams in IP networks'. Proc. Electronics, Robotics and Automotive Mechanics Conf. (CERMA), September 2006, vol. 1, pp. 178–183

[29] RIES M., NEMETHOVA O., RUPP M.: 'Video quality estimation for mobile H.264/AVC video streaming', *J. Commun.*, 2008, **3**, (1), pp. 41–50

[30] KHAN A., SUN L., IFEACHOR E.: 'Content-based video quality prediction for MPEG4 video streaming over wireless networks', *J. Multimedia*, 2009, **4**, (4), pp. 228–239

[31] JANG R.: 'ANFIS: adaptive-network-based fuzzy inference systems', *IEEE Trans. Syst. Man Cybern.*, 1993, **23**, (3), pp. 665–685

[32] Ffmpeg, http://sourceforge.net/projects/ffmpeg

[33] JM H.264 Software, http://iphome.hhi.de/suehring/tml/

[34] NS2, http://www.isi.edu/nsnam/ns/

[35] KLAUE J., TATHKE B., WOLISZ A.: 'Evalvid – a framework for video transmission and quality evaluation'. Proc. 13th Int. Conf. on Modelling Techniques and Tools for Computer Performance Evaluation, Urbana, Illinois, USA, 2003, pp. 255–272

[36] Enhanced UMTS Radio Access Network Extensions for ns-2 (e.u.r.a.n.e); http://eurane.ti-wmc.nl/eurane/

[37] KHAN A., SUN L., FAJARDO J., LIBERAL F., IFEACHOR E.: 'Video quality prediction model for H.264 video over UMTS networks and their application in mobile video streaming'. IEEE ICC, Cape Town, South Africa (accepted for publication), 23–27 May 2010

[38] SNEDECOR G.W., COCHRAN W.G.: 'Statistical methods' (Iowa State University Press, 1989, 8th edn.)

[39] HOCHBERG Y., TAMHANE A.C.: 'Multiple comparison procedures' (Wiley, 1987)

[40] ITU: 'Subjective video quality assessment methods for multimedia applications'. ITU-T Recommendation P.910, Geneva, September 1999

[41] http://www.tech.plym.ac.uk/staff/akhan/mostest/default4.htm